DANNY SIU

https://dannysiudata.com
(817) 320-7283
dannysiu@berkeley.edu

EDUCATION

Berkeley, CA
Aug. 2017 - May 2021

University of California at Berkeley
Bachelor of Arts in Data Science; Domain Emphasis in Linguistics; Minor in Computer Science
National Merit Scholar; Valedictorian of Lamar High School

Coursework:

- Algorithms
- Data Structures
- Computer Security
- Machine Learning and Data Analytics
- Artificial Intelligence (AI)
- Syntax
- Natural Language Processing (NLP)
- Database Systems
- Computer Architecture
- Data Mining and Analytics
- Lexical Semantics
- Sociolinguistics

EXPERIENCE

Berkeley, CA
Feb. 2019 - Aug. 2019

University of California at Berkeley, Cohen Group - Research Intern

- Launched efforts to replace a computationally expensive WRF-CHEM climate model run with machine learning model predictions.
- Implemented preparations to produce higher resolution measurements of NO2 levels than measured by NASA's Aura satellite.
- Aligned 47K WRF-CHEM pixels with Aura's OMI pixels to enable better data correlation.
- Created easily visualized mapping of geolocation data measurements.

Austin, TX
Jun. 2016 - Aug. 2016

University of Texas at Austin - Summer Student Researcher

- Optimized catalysts for improved efficiency in energy production.
- Created molecular visualization movies of catalyst binding to communicate research.
- Computationally simulated carbon monoxide binding on alloys of gold-palladium nanoparticles during oxidation-reduction reactions.

PROJECTS

More on my website: **https://dannysiudata.com**

Python
(PyTorch, NumPy)

Neural Sequence Labeling

- Implemented and trained an LSTM for POS-tagging by using mini-batch stochastic gradient descent on pre-trained GloVe embeddings.
- Improved model accuracy by implementing a bi-directional LSTM and dropout layer, and loading the 500,000 most common 300-dimensional word embeddings.

Python
(Scikit-Learn, SciPy, NumPy)

Movie Review Sentiment Classifier

- Created a binary sentiment classifier for movie reviews using featurization of bag-of-words, bigrams, trigrams, sentiment dictionaries, and neutral word removal.

Java

PokeMan Game

- Developed a multi-player game featuring Pokemon characters in randomized worlds.
- Implemented game saving using Git, menus, and a live heads-up display (HUD).

Python, Excel

Study of Sociolinguistic Variation for Affirmation and Declination Responses

- Explored the sociolinguistic variation in "yes" and "no" response variants for English speakers.
- Collected auditory linguistic data over the span of two weeks to determine if there was any significant correlation between variant usage and social factors of gender, race, formality of relationship, and situational context.

SKILLS

Programming: Python, Java, SQL, C, Unix/Linux, Git, PyTorch, HTML, CSS, VASP

Data Science: Data modeling (Classification, Prediction, Clustering, Gradient descent), Data visualization, Machine learning, Feature engineering, Data analysis, Data mining

NLP: Neural networks (RNNs, CNNs), LSTMs, Hidden Markov Models, Sentiment analysis, Dependency parsing, Text embeddings (Word2vec, GloVe, FastText), One-hot encoding